Employee Training Impact Analysis

Analysing the Impact of a New Training Program on Employee Performance using paired sample t-test

Hypothesis

Null Hypothesis : H_0 -> There is no significant difference between the evaluation scores before and after the training

Alternate Hypothesis : H_1 -> There is a significant difference between the evaluation scores before and after the training

Datset

	employee_id	department	region	education	gender	recruitment_channel	no_of_trainings	age	previous_year_rating	length_of_service	KPIs_met >80%	awards_won?	avg_training_score	last_evaluation_score
0	3	Operations	region_22	Master's & above	f	other	1	38	5.0	4	1	0	67	0.73
1	6	Operations	region_15	Bachelor's	f	other	1	38	3.0	10	0	0	50	0.76
2	11	Sales & Marketing	region_4	Master's & above	f	sourcing	1	39	4.0	12	1	0	72	0.58
3	13	Operations	region_17	Bachelor's	m	other	1	38	1.0	5	1	0	47	0.39
4	16	Operations	region_4	Master's & above	m	sourcing	1	40	5.0	5	0	0	53	0.58
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
23485	78280	Analytics	region_7	Bachelor's	m	sourcing	2	30	NaN	1	0	0	51	1.00
23486	78283	Procurement	region_12	Bachelor's	f	sourcing	1	28	3.0	4	0	0	57	0.84
23487	78284	Technology	region_22	Bachelor's	m	referred	1	33	3.0	5	1	1	82	0.62
23488	78293	Sales & Marketing	region_27	Bachelor's	f	other	1	26	3.0	3	1	0	64	0.87
23489	78295	Operations	region_14	Bachelor's	f	sourcing	1	36	3.0	5	0	0	48	0.62

Performing Z-score Normalization

We are performing Z-score normalization of ***previous_year_rating*** to match the scale of ***last_evaluation_score***

previous_year_rating
5.0
3.0
4.0
1.0
5.0
...

new_prev
1.308394
-0.276078
0.516158
-1.860551
1.308394
...

Variable Selection

new_prev : Performance score of the employees before the training

last_evaluation_score : Performance score of the employees after the training

Group Identification

```
data['department'].value_counts()
```

```
department
Sales & Marketing    6450
Operations           4234
Technology           2692
Procurement          2668
Analytics            2057
HR                    976
Finance               958
Legal                 397
R&D                   387
```

- We identify that there are 9 different departments in the company and all the employees of every department are undergoing a training program respective of their department

Performing t-test on Analytics Department

```
analytics=data[(data.department=="Analytics") & (data.no_of_trainings==1)]
analytics_bef=analytics['new_prev'].sample(n=29)
analytics_aft=analytics['last_evaluation_score'].sample(n=29)
```

```
tana,p_ana=stats.ttest_rel(analytics_bef,analytics_aft)
```

```
print("tstatistic value is : ",tstatistic)
print("p_value  is : ",p_ana)
```

```
tstatistic value is :  -3.788769516266795
p_value  is :  0.07249861757831698
```

- We are selecting random 29 employees from the ***Analytics*** Department and performing paired sample t-test on those samples

Findings on Analytics Department

```
tstatistic value is :  -3.788769516266795
p_value  is :  0.07249861757831698
```

```
if p_ana<0.05:
    print("We reject the null hypothesis (i.e) There is a significant difference between the evaluation scores before and after the training")
else:
    print("We fail to reject the null hypothesis (i.e) There is no significant difference between the evaluation scores before and after the training")
```

```
We fail to reject the null hypothesis (i.e) There is no significant difference between the evaluation scores before and after the training
```

- Since the p-value is greater than the significance level 0.05, we fail to reject the null hypothesis and identify that there is no significant difference between the evaluation scores before and after the training (i.e) ***The Training does not lead to higher last evaluation scores***

Performing t-test on Finance Department

```
finance=data[(data.department=="Finance") & (data.no_of_trainings==1)]
finance_bef=finance['new_prev'].sample(n=29)
finance_aft=finance['last_evaluation_score'].sample(n=29)
```

```
tstatistic,p_fin=stats.ttest_rel(finance_bef,finance_aft)
```

```
print("tstatistic value is : ",tstatistic)
print("p_value  is : ",p_fin)
```

```
tstatistic value is :  -3.4070411361615
p_value  is :  0.002005758534820048
```

- We are selecting random 29 employees from the ***Finance*** Department and performing paired sample t-test on those samples

Findings on Finance Department

```
tstatistic value is :  -3.4070411361615
p_value  is :  0.002005758534820048
```

```
if p_fin<0.05:
    print("We reject the null hypothesis (i.e) There is a significant difference between the evaluation scores before and after the training")
else:
    print("We fail to reject the null hypothesis (i.e) There is no significant difference between the evaluation scores before and after the training")
```

```
We reject the null hypothesis (i.e) There is a significant difference between the evaluation scores before and after the training
```

- Since the p-value is less than the significance level 0.05, we are rejecting the null hypothesis and identify that there is a significant difference between the evaluation scores before and after the training (i.e) ***The Training leads to higher last evaluation scores***

Similarly performing t-test on other Departments

```
tstatistic,p_leg=stats.ttest_rel(Legal_bef,Legal_aft)
```

```
tstatistic,p_ope=stats.ttest_rel(Operations_bef,Operations_aft)
```

```
tstatistic,p_hr=stats.ttest_rel(HR_bef,HR_aft)
```

```
tstatistic,p_sal=stats.ttest_rel(Sales_bef,Sales_aft)
```

```
tstatistic,p_rd=stats.ttest_rel(Rd_bef,Rd_aft)
```

```
tstatistic,p_pro=stats.ttest_rel(Procurement_bef,Procurement_aft)
```

```
tstatistic,p_tec=stats.ttest_rel(Technology_bef,Technology_aft)
```

- We are selecting random 29 employees from the each Department and performing paired sample t-test on those samples

Findings on other Departments

Department	tstatistic	p_values	significance_level
Legal	-4.648181	0.000073	0.05
Operations	-2.562763	0.016049	0.05
HR	-2.469115	0.019910	0.05
Sales & Marketing	-3.732777	0.000856	0.05
R&D	-3.495739	0.001594	0.05
Procurement	-3.533138	0.001446	0.05
Technology	-3.934058	0.000501	0.05

- Since the p-value is less than the significance level 0.05 on remaining department's t-test, we are rejecting the null hypothesis and identify that there is a significant difference between the evaluation scores before and after the training on remaining departments (i.e) ***The Training leads to higher last evaluation scores***

Conclusion

	Department	tstatistic	p_values	significance_level
0	Analytics	-1.866350	0.072499	0.05
1	Finance	-4.713894	0.000061	0.05
2	Legal	-4.648181	0.000073	0.05
3	Operations	-2.562763	0.016049	0.05
4	HR	-2.469115	0.019910	0.05
5	Sales & Marketing	-3.732777	0.000856	0.05
6	R&D	-3.495739	0.001594	0.05
7	Procurement	-3.533138	0.001446	0.05
8	Technology	-3.934058	0.000501	0.05

- In conclusion, we can say that except the ***Analytics*** department there is a significant difference between the evaluation scores before and after the training
- In other words, the Training program is useful for all the departments except the Analytics department

Thank You